Numerische Mathematik manuscript No.
(will be inserted by the editor)

A Variational Principle for Adaptive Approximation of Ordinary Differential Equations

Kyoung-Sook Moon[1], Anders Szepessy[1 2], Raúl Tempone[1], Georgios E. Zouraris[3]

[1] Institutionen för Numerisk Analys och Datalogi, Kungl Tekniska Högskolan, S–100 44 Stockholm; e-mail: moon@nada.kth.se, rtempone@nada.kth.se
[2] Matematiska Institutionen, Kungl Tekniska Högskolan, S–100 44 Stockholm; e-mail: szepessy@math.kth.se
[3] Department of Mathematics, University of the Aegean, GR–832 00 Karlovassi, Samos, Greece, and IACM, FO.R.T.H, GR–711 10 Heraklion, Greece; e-mail: zouraris@aegean.gr

Received: date / Revised version: date

Summary A variational principle, inspired by optimal control, yields a simple derivation of an error representation, global error $= \sum$ local error $\cdot$ weight, for general approximation of functions of solutions to ordinary differential equations. This error representation is then approximated by a sum of computable error indicators, to obtain a useful global error indicator for adaptive mesh refinements. A uniqueness formulation is provided for desirable error representations of adaptive algorithms.

Mathematics Subject Classification (1991): 65L70, 65G50

Key words a posteriori error estimate, local error, ordinary differential equations, variational principle, roundoff error.

1 Introduction to Error Expansions for ODE

This paper derives an error expansion, for approximation of ordinary differential equations, of the form

$$\text{global error} = \sum_{\text{time steps}} \text{local error} \cdot \text{weight} + \text{higher order error}. \quad (1.1)$$

Such error estimates for differential equations can be derived by the classical error equation and linearization, cf. [13], [8], [12], by Galerkin orthogonality using either local problems or the residual, cf. [1], [9], and by a variational principle, following Alekseev [2] and Gröbner [11]. The variational principle was introduced to derive an error representation for perturbation errors of differential equations based on the residual of the perturbation. Our analysis applies the variational principle to error analysis based on local errors and gives a simple and precise derivation of the fundamental property that the global error is a weighted sum of the local errors

$$\text{global error} = \sum_{\text{time steps}} \text{local error} \cdot \text{weight}. \tag{1.2}$$

However in (1.2) both the true local errors and the weights are non computable, therefore we transfer the representation (1.2) to an asymptotic expansion (1.1) with computable approximations of the local errors and the weights, using standard estimates. The adaptive algorithm analyzed and tested in [18] is based on the leading order term in the expansion (1.1). Computations including similar asymptotic global error control are well known, cf. [13], [19]. Section 3 shows that the classical derivation of (1.1) based on the error equation does not include (1.2), since the global error is polluted by the linearization error between the exact and approximate solution.

Consider a solution $X : [0,T] \to \mathbb{R}^d$ of a differential equation, with flux $a : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$,

$$\begin{aligned} \frac{dX}{dt}(t) &= a(t, X(t)), \quad 0 < t \leq T, \\ X(0) &= X_0, \end{aligned} \tag{1.3}$$

and an approximation $\overline{X}$ of (1.3) by any numerical method, satisfying the same initial condition

$$\overline{X}(0) = X(0) = X_0 \tag{1.4}$$

with time steps

$$0 = t_0 < \cdots < t_N = T.$$

The next section derives estimates of the form (1.1) for the global error

$$g(X(T)) - g(\overline{X}(T)) \tag{1.5}$$

with a given general function $g : \mathbb{R}^d \to \mathbb{R}$. The function g is therefore included in the data of the problem, which the user specifies as in optimal control problems; i.e. the user provides the information to

approximate the value of the objective function g in the algorithm. One example is to find the value of one component of the solution at the final time, e.g. $g(x) = x_1$. Note that the apparently more general approximation of $\int_0^T h(X(t), t)dt + g(X(T))$, for a given function $h : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$, is a particular case of (1.5) by extending the system (1.3) with the additional equation $dX_{d+1}(t)/dt = h(X(t), t)$ and the objective $g(x) + x_{d+1}$.

The estimates (1.1)-(1.2) will use the local error e defined by

$$e(t_n) \equiv \widetilde{X}(t_n) - \overline{X}(t_n), \tag{1.6}$$

where the local exact solution $\widetilde{X}$ satisfies, for each time step $(t_{n-1}, t_n]$,

$$\begin{aligned} \frac{d\widetilde{X}}{dt}(t) &= a(t, \widetilde{X}(t)), \quad t_{n-1} < t \le t_n, \\ \widetilde{X}(t_{n-1}) &= \overline{X}(t_{n-1}). \end{aligned} \tag{1.7}$$

Theorems 2.1 and 2.3-2.4, below, prove error estimates of the form (1.2) and (1.1), respectively. These results use the weight function, which solves a certain linear backwards dual (or adjoint) problem, obtained by linearizing the forward problem (1.3) around the solution. In an adaptive time stepping method with a given bound on the global error, the number of time steps are minimized by choosing for all time steps

$$\big|\text{local error} \cdot \text{weight}\big| = \text{constant}. \tag{1.8}$$

Therefore the weights need to be determined to find the optimal mesh. It is possible to approximate the weight with lower accuracy and a coarser mesh than that used for the approximate solution $\overline{X}$. When the solution is well resolved, the work to determine the weight can therefore be smaller than the work to solve the differential equation (1.3). However, to solve the dual problem requires the storage of the approximate solution on the coarser mesh. This additional storage is clearly a drawback. On the other hand many computer programs for the numerical solution of ordinary differential equations store the solution at all time levels for other reasons, e.g. for post processing. The use of dual functions is standard in optimal control theory and also well known for adaptive mesh control for ordinary and partial differential equations, see [3], [4], [6], [10], [15], [16], [20].

In conclusion, the main results are:

- an easy derivation of the fundamental property "global error = $\sum$ local error $\cdot$ weight" which can be used, e.g., in introductory numerical analysis courses;

– analysis of an error expansion useful for adaptive mesh refinements method based on global error estimation, including a formulation for uniqueness of desirable error representations for adaptive refinements.

The outline of the paper is: Section 2 and 2.1-2.3 prove by a variational principle an error representation (1.2) and an error expansion (1.1). Section 2.4 treats roundoff errors. Section 2.5 provides a formulation of uniqueness for adaptive error representations. Finally Section 3 compares the derivation in Section 2 to the well known alternatives to obtain (1.1) by the error equation and by the residual.

2 The Variational Principle

Let $X(s;t,y)$ denote the solution of (1.3) at time s, which at time t takes the value y, i.e.

$$\begin{aligned} &\frac{dX}{ds}(s;t,y) = a(s,X(s;t,y)), \quad t < s \leq T, \\ &X(t;t,y) = y. \end{aligned} \tag{2.1}$$

Define, for the given function $g : \mathbb{R}^d \to \mathbb{R}$ in (1.5), the function $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ by

$$u(t,y) \equiv g(X(T;t,y)), \quad t < T, \tag{2.2}$$

provided the differential equation (2.1) on (t,T) has a unique solution for all initial data y in $\mathbb{R}^d$ and all $t \in (0,T)$. In the following theorem, the global approximation error for differential equations is represented in terms of the local errors and their weights, depending on the first variation of u. The generalization to partial differential equations is then possible with some convenient assumptions, see [17].

Theorem 2.1 *Assume that* (2.1) *has a unique continuous solution* X *for all initial data* $y \in \mathbb{R}^d$ *and that the flux* $a(t,x)$ *is differentiable in* x*, for all* $t \in (0,T)$*. Let* $e(t_n) \equiv \widetilde{X}(t_n) - \overline{X}(t_n)$ *denote the local error of an approximation,* $\overline{X}$*, of* (1.3)*, satisfying* (1.4) *and* (1.7)*. Then, for any differentiable function* $g : \mathbb{R}^d \to \mathbb{R}$*, the function* u *is well defined by* (2.2) *and the global error is a weighted sum of the local error with the representation*

$$\begin{aligned} &g(X(T)) - g(\overline{X}(T)) \\ &\qquad = \sum_{n=1}^{N} \left(e(t_n), \int_0^1 \Psi\Big(t_n, \overline{X}(t_n) + se(t_n)\Big) ds \right), \end{aligned} \tag{2.3}$$

where $(\cdot,\cdot)$ *is the standard scalar product on* $\mathbb{R}^d$ *and* $\Psi(t,y) \equiv \Psi_X(t) \in \mathbb{R}^d$ *is the first variation of* u *in the sense that for all* $w \in \mathbb{R}^d$ *and all sufficiently small* $\delta > 0$

$$u(t, y + \delta w) - u(t, y) = (\Psi_X(t), \delta w) + o(\delta).$$

The weight function Ψ_X *satisfies, for* $t < s < T$*, the dual equation*

$$\begin{aligned} -\frac{d\Psi_X}{ds}(s) &= (a')^*(s, X(s;t,y))\, \Psi_X(s), \\ \Psi_X(T) &= g'(X(T;t,y)), \end{aligned} \tag{2.4}$$

where $(a')^*(s,x)$ *is the transpose of the Jacobian matrix* $a'(s,x) \equiv \{\frac{\partial a_i}{\partial x_j}(s,x)\} \in \mathbb{R}^{d\times d}$*, and* X *solves* (2.1)*.*

Proof. By the construction (2.2), the function u is constant along the characteristics $\widetilde{X}$, i.e. for all t and τ in $[t_{n-1}, t_n]$

$$u(t, \widetilde{X}(t)) = u(\tau, \widetilde{X}(\tau)). \tag{2.5}$$

Therefore the initial condition for the local problem (1.7) shows that

$$u(t_n, \widetilde{X}(t_n)) = u(t_{n-1}, \widetilde{X}(t_{n-1})) = u(t_{n-1}, \overline{X}(t_{n-1})), \quad n = 1, \ldots, N,$$

and consequently the initial condition (1.4) and (2.2) imply that

$$\begin{aligned} \sum_{n=1}^{N} (u(t_n, \widetilde{X}(t_n)) - u(t_n, \overline{X}(t_n))) &= u(0, \overline{X}(0)) - u(T, \overline{X}(T)) \\ &= u(0, X(0)) - u(T, \overline{X}(T)) \\ &= g(X(T)) - g(\overline{X}(T)). \end{aligned} \tag{2.6}$$

The function $U : [0,1] \to \mathbb{R}$, defined by

$$U(s) = u(t_n, s\widetilde{X}(t_n) + (1-s)\overline{X}(t_n)),$$

and the equality

$$U(1) - U(0) = \int_0^1 U'(s) ds$$

show that each term in the sum (2.6) can be written

$$u(t_n, \widetilde{X}(t_n)) - u(t_n, \overline{X}(t_n)) = \left(e(t_n), \int_0^1 \Psi(t_n, \overline{X}(t_n) + se(t_n)) ds \right),$$

which proves (2.3). The first variation $\partial X(s;t,y)/\partial y$ exists, since $a(s,x)$ is differentiable in x. The combination of the existence of the first variation of X and the assumption that g is differentiable, implies

by (2.1)-(2.2) that Ψ exists. Finally, to verify that Ψ satisfies the dual equation (2.4), observe that, for any $w \in \mathbb{R}^d$ and $\delta \to 0$, two solutions X^1 and X^2 of (2.1), with initial data $X^1(t) = y \in \mathbb{R}^d$ and $X^2(t) = y + \delta w$ satisfy

$$\begin{aligned}
0 &= \frac{d}{dt}\left(u(t, X^2(t)) - u(t, X^1(t))\right) \\
&= \frac{d}{dt}\left(\Psi_{X^1}, X^2(t) - X^1(t)\right) + o(\delta) \\
&= \left(\frac{d}{dt}\Psi_{X^1}, X^2(t) - X^1(t)\right) + \left(\Psi_{X^1}, \frac{d}{dt}X^2(t) - \frac{d}{dt}X^1(t)\right) + o(\delta) \\
&= \left(\delta\frac{d}{dt}\Psi_{X^1}, w\right) + \left(\delta\Psi_{X^1}, a'(t, X^1(t))w\right) + o(\delta),
\end{aligned}$$

which proves (2.4) in the limit $\delta \to 0$. □

Our next goal is to construct error expansions useful for adaptive methods. The starting point for the adaptive method in [18] is an asymptotic expansion of the representation (2.3) with leading order term in computable form. Such error expansions are well known and derived e.g. in [13], [12]. The motivation to derive this expansion again is that ours has somewhat sharper higher order terms; Section 3 presents a comparison between these global error expansions derived by the variational principle, the residual with Galerkin orthogonality and the error equation. However, the higher order terms are not used in an essential way in the adaptive algorithm in [18].

The error expansion is based on an approximation $\overline{\Psi}$ of the weight Ψ and an approximation $\bar{e}$ of the local error e

$$\begin{aligned}
&\sum_{n=1}^{N}\left(e(t_n), \Psi\right) - \sum_{n=1}^{N}\left(\bar{e}(t_n), \overline{\Psi}\right) \\
&\qquad = \sum_{n=1}^{N}\left(e(t_n) - \bar{e}(t_n), \overline{\Psi}\right) + \sum_{n=1}^{N}\left(e(t_n), \Psi - \overline{\Psi}\right).
\end{aligned} \tag{2.7}$$

2.1 Approximation of the Weight

The averaged weight function Ψ in Theorem 2.1, which is needed to determine the optimal step size in an adaptive method, can be computed by approximating (2.1) and (2.4). Therefore, any p-th order

accurate approximation $(\overline{X},\overline{\Psi})$ of (X,Ψ), which solves the systems of differential equations (1.3) and (2.4), satisfies

$$\left|\overline{\Psi}(t_n) - \Psi(t_n,\overline{X}(t_n))\right| = \mathcal{O}((\max \Delta t)^p) + \mathcal{O}\left(\frac{\epsilon}{\min \Delta t}\right), \qquad (2.8)$$

where ϵ is the machine roundoff unit and $\Delta t_n = t_n - t_{n-1}$ with $\max \Delta t \equiv \max_n \Delta t_n$ and $\min \Delta t \equiv \min_n \Delta t_n$. The effect of roundoff errors is neglected in the formulation of Theorems 2.2 and 2.3 below. Instead a remark in the end of the section includes roundoff errors in the error estimation.

A natural choice of approximation $\overline{\Psi}$, for a p-th order one step method $\overline{X}$ written in the form

$$\overline{X}(t_n) = A(\overline{X}(t_{n-1}), \Delta t_n), \qquad (2.9)$$

is

$$\begin{aligned} \overline{\Psi}_i(t_{n-1}) &= \sum_{j=1}^{d} \partial_{x_i} A_j(\overline{X}(t_{n-1}), \Delta t_n)\overline{\Psi}_j(t_n), \\ \overline{\Psi}_i(T) &= \partial_{x_i} g(\overline{X}(T)), \end{aligned} \qquad (2.10)$$

which yields a p-th order accurate approximation $(\overline{X},\overline{\Psi})$ of (X,Ψ) and satisfies

$$\overline{\Psi}_i(t_{n-1}) = \partial_{x_i} g\left(\overline{X}(T;\overline{X}(t_{n-1}) = x)\right). \qquad (2.11)$$

The relation (2.11) is the discrete version of the fact that $\Psi(t)$ is the first variation of $g(X(T))$ with respect to variation in the location of the path $X(t)$ at time t, and (2.11) holds precisely when $\overline{\Psi}$ is defined by (2.10). The Jacobian matrix $\partial_{x_i} A_j(\overline{X}(t_{n-1}), \Delta t_n) = \frac{\partial \overline{X}_j(t_n)}{\partial \overline{X}_i(t_{n-1})}$ can be approximated by numerical differentiation of $\overline{X}(t_n)$ with respect to $\overline{X}(t_{n-1})$, or alternatively the Jacobian can be evaluated explicitly for each method, e.g. to preserve a sparse structure. To conclude, we have the error estimate

Theorem 2.2 *Suppose that* (2.8) *and the assumptions of Theorem 2.1 hold. Let* $\partial_{xx}u(t,x)$ *in* (2.2) *be uniformly bounded for* $(t,x) \in [0,T]\times\mathbb{R}^d$. *Then the global approximation error for the differential equation* (1.3) *satisfies the estimate*

$$\begin{aligned} &g(X(T)) - g(\overline{X}(T)) \\ &\qquad = \sum_{n=1}^{N}\Big(e(t_n), \overline{\Psi}(t_n) + \mathcal{O}(|e(t_n)|) + \mathcal{O}((\max \Delta t)^p)\Big) \end{aligned} \qquad (2.12)$$

where $e(t_n) \equiv \widetilde{X}(t_n) - \overline{X}(t_n)$ *is the local error and* $(\overline{X},\overline{\Psi})$ *is a p-th order accurate approximation of the system* (1.3) *and* (2.4).

Proof. A combination of Theorem 2.1, (2.8) and the boundedness of $\partial_{xx}u \equiv \partial_x \Psi$ implies that

$$\begin{aligned}
&\Psi(t_n, \overline{X}(t_n) + se(t_n)) - \overline{\Psi}(t_n) \\
&= \big(\Psi(t_n, \overline{X}(t_n) + se(t_n)) - \Psi(t_n, \overline{X})\big) + \big(\Psi(t_n, \overline{X}) - \overline{\Psi}(t_n)\big) \\
&= \mathcal{O}(e(t_n)) + \mathcal{O}((\max \Delta t)^p),
\end{aligned}$$

which proves (2.12). □

2.2 Approximation of the Local Error

The next step necessary to derive an error estimate based on computable quantities is to approximate the local error $e = \widetilde{X} - \overline{X}$ by replacing the exact local solution $\widetilde{X}$ by an approximation $\overline{\overline{X}}$ of higher accuracy than $\overline{X}$, i.e., with smaller time steps or a higher order method in a higher precision. For smooth solutions X, the existence of the limits

$$\begin{aligned}
&\lim_{\Delta t \to 0} (\Delta t_n)^{-(p+1)} \big(\widetilde{X}(t_n) - \overline{X}(t_n)\big), \\
&\lim_{\Delta t \to 0} (\Delta t_n)^{-(q+1)} \big(\widetilde{X}(t_n) - \overline{\overline{X}}(t_n)\big),
\end{aligned} \tag{2.13}$$

determines by Richardson extrapolation a constant γ, for $q \geq p$ cf. [7], such that

$$e(t_n) = \widetilde{X}(t_n) - \overline{X}(t_n) = \gamma \left(\overline{\overline{X}}(t_n) - \overline{X}(t_n)\right) + o(\Delta t_n^{p+1}). \tag{2.14}$$

For instance there holds: $\gamma = 2^p/(2^p - 1)$ for $\overline{\overline{X}}$ computed with the half mesh size and $q = p$; and $\gamma = 1$ for $\overline{\overline{X}}$ computed with a higher order method $q > p$, see [12]. Let $\Delta t(t) \equiv \Delta t_n$, $t_{n-1} < t \leq t_n$. The replacement of the exact local error with this approximate local error leads to

Theorem 2.3 *Suppose that the limits* (2.13) *exist and that the assumptions of Theorems 2.1 and 2.2 hold. Then the global approximation error for the differential equation* (1.3) *satisfies the estimate*

$$g(X(T)) - g(\overline{X}(T)) = \sum_{n=1}^{N} \big(\bar{e}(t_n), \overline{\Psi}(t_n)\big) + \int_0^T o(\Delta t^p(t))dt \tag{2.15}$$

where $\bar{e}(t_n) \equiv \gamma \left(\overline{\overline{X}}(t_n) - \overline{X}(t_n)\right)$ *is the approximation of the local error in* (2.14) *and* $(\overline{X}, \overline{\Psi})$ *is a p-th order accurate approximation of the system* (1.3) *and* (2.4).

Proof. By substituting (2.14) in (2.12), we obtain

$$g(X(T)) - g(\overline{X}(T)) = \sum_{n=1}^{N} \big(\bar{e}(t_n), \overline{\Psi}(t_n)\big) + \mathcal{E}$$

where

$$\begin{aligned}\mathcal{E} &\equiv \sum_{n=1}^{N} \big[\ o(\Delta t^{p+1}) \ + \ \mathcal{O}(\Delta t^{p+1}) \cdot \big(\Delta t^{p+1} + (\max \Delta t)^p\big)\ \big] \\ &= \int_0^T o(\Delta t^p) dt \\ &\leq o((\max \Delta t)^p),\end{aligned}$$

which proves the theorem. □

2.3 An Alternative Error Expansion

Let us present an alternative to the error expansion (2.15), using Richardson extrapolation. Assume that $\overline{X}^h$ and $\overline{X}^H$ are approximations based on (2.9), where the step sizes $\Delta_h t(t)$ and $\Delta_H t(t)$ satisfy

$$\begin{aligned}&\frac{\Delta_h t(t)}{\Delta_H t(t)} \text{ is independent of } t, \\ &\max_t \Delta_h t(t) = h, \\ &\max_t \Delta_H t(t) = H.\end{aligned} \tag{2.16}$$

Then the convergence assumption

$$\begin{aligned}g(X(T)) - g(\overline{X}^h(T)) &= c\ h^p + \alpha_h, \\ \alpha_h &= o(h^p),\end{aligned} \tag{2.17}$$

is meaningful for two positive constants c and p. An attractive alternative to the approximation (2.15) is to use (2.17) and apply Richardson extrapolation directly to the p-th order accurate approximations $\overline{X}^h$ and $\overline{X}^H$ of X to obtain

Theorem 2.4 *Suppose that* (2.16), (2.17) *and the assumptions of Theorem 2.1 hold. Let the two p-th order accurate approximations* $\overline{X}^h$ *and* $\overline{X}^H$ *of* X*, be defined by*

$$\overline{X}^h(t_n) = A^h(\overline{X}^h(t_{n-1})), \quad \overline{X}^H(t_n) = A^H(\overline{X}^H(t_{n-1})) \tag{2.18}$$

following (2.9)*. Then the global approximation error for the differential equation* (1.3) *satisfies the estimate*

$$g(X(T)) - g(\overline{X}(T)) = \alpha + \frac{1}{(\frac{H}{h})^p - 1} \sum_{n=1}^{N} \left(A^h(\overline{X}^H(t_{n-1})) - A^H(\overline{X}^H(t_{n-1})), \overline{\Phi}(t_n) \right), \tag{2.19}$$

where

$$\alpha \equiv \alpha_h + \frac{\alpha_h - \alpha_H}{(\frac{H}{h})^p - 1} = o(h^p) \tag{2.20}$$

and the weight function $\overline{\Phi}(t_n) \in \mathbb{R}^d$ *is defined for* $n = N, \ldots, 1$ *and* $i = 1, \ldots, d$ *by the recursive equation*

$$\begin{aligned} &\overline{\Phi}_i(T) = \int_0^1 \partial_{x_i} g \left(s\overline{X}^h(T) + (1-s)\overline{X}^H(T) \right) ds, \\ &\overline{\Phi}_i(t_{n-1}) \\ &\quad = \left(\int_0^1 \partial_{x_i} A^h \left(s\overline{X}^h(t_{n-1}) + (1-s)\overline{X}^H(t_{n-1}) \right) ds, \overline{\Phi}(t_n) \right). \end{aligned} \tag{2.21}$$

Proof. Using Richardson extrapolation, we get

$$\begin{aligned} g(X(T)) - g(\overline{X}(T)) \\ = \frac{1}{(\frac{H}{h})^p - 1} \left(g(\overline{X}^h(T)) - g(\overline{X}^H(T)) \right) + o(h^p). \end{aligned} \tag{2.22}$$

Therefore it is sufficient to prove that the computable quantity,

$$g(\overline{X}^h(T)) - g(\overline{X}^H(T)),$$

has the representation

$$\begin{aligned} g(\overline{X}^h(T)) - g(\overline{X}^H(T)) \\ = \sum_{n=1}^{N} \left(A^h(\overline{X}^H(t_{n-1})) - A^H(\overline{X}^H(t_{n-1})), \overline{\Phi}(t_n) \right) \end{aligned} \tag{2.23}$$

which proves (2.19)-(2.20) together with (2.22). The initial conditions in (2.21) and (1.4) and telescoping cancellation show

$$\begin{aligned} g(\overline{X}^h(T)) - g(\overline{X}^H(T)) &= \left(\overline{X}^h(t_N) - \overline{X}^H(t_N), \overline{\Phi}(t_N) \right) \\ &= \sum_{n=1}^{N} \Big[\left(\overline{X}^h(t_n) - \overline{X}^H(t_n), \overline{\Phi}(t_n) \right) \\ &\qquad - \left(\overline{X}^h(t_{n-1}) - \overline{X}^H(t_{n-1}), \overline{\Phi}(t_{n-1}) \right) \Big]. \end{aligned} \tag{2.24}$$

By the definitions (2.18), the right hand side of (2.24) can be separated into three parts

$$\Big(A^h(\overline{X}^h(t_{n-1})) - A^h(\overline{X}^H(t_{n-1})), \overline{\Phi}(t_n)\Big),$$
$$\Big(A^h(\overline{X}^H(t_{n-1})) - A^H(\overline{X}^H(t_{n-1})), \overline{\Phi}(t_n)\Big),$$
$$-\Big(\overline{X}^h(t_{n-1}) - \overline{X}^H(t_{n-1}), \overline{\Phi}(t_{n-1})\Big),$$

where the first and the last parts are canceled out, since the first term can be written

$$\Big(A^h(\overline{X}^h(t_{n-1})) - A^h(\overline{X}^H(t_{n-1})), \overline{\Phi}(t_n)\Big)$$
$$= \Big(\overline{X}^h(t_{n-1}) - \overline{X}^H(t_{n-1}), \sum_{j=1}^{d} \Lambda_{\cdot j}\, \overline{\Phi}_j(t_n)\Big)$$

with

$$\Lambda_{ij} \equiv \int_0^1 \partial_{x_i} A_j^h\Big(s\overline{X}^h(t_{n-1}) + (1-s)\overline{X}^H(t_{n-1})\Big)ds,$$

so that

$$\sum_j \Lambda_{ij}\overline{\Phi}_j(t_n) = \overline{\Phi}_i(t_{n-1})$$

by (2.21). Consequently (2.23) holds. □

2.4 Approximation of Roundoff Error

Theorems 2.2 and 2.3 can be modified to include error caused by roundoff due to finite precision arithmetic. Let $\hat{e}(t_n)$ be the local roundoff error in one step of the method $\overline{X}$, i.e.,

$$\hat{e}(t_n) \equiv \widehat{\overline{X}}(t_n) - \overline{X}(t_n),$$

where $\widehat{\overline{X}}(t_n)$ is the exact arithmetic version of one step of the method $\overline{X}$, with the same initial data $\overline{X}(t_{n-1})$ at time t_{n-1}. Here, $\overline{X}$ is computed in finite precision arithmetic. Theorem 2.1 then implies that the part of the global error due to roundoff is the following weighted sum of the local roundoff error

$$\sum_{n=1}^{N}\left(\hat{e}(t_n), \int_0^1 \Psi\left(t_n, \overline{X}(t_n) + se(t_n)\right)ds\right).$$

If the local roundoff error $\hat{e}(t_n)$ in one step dominates the local discretization error, i.e., if

$$|\hat{e}(t_n)| \geq \left|\gamma\big(\overline{\overline{X}}(t_n) - \overline{X}(t_n)\big)\right|, \tag{2.25}$$

then the refinement of the time step will not decrease the approximation error; instead a higher precision is needed. The local roundoff error $\hat{e}$ can be estimated by approximating $\overline{\widehat{X}}$ with higher precision than $\overline{X}$. Alternatively, the approximation

$$fl\big(fl(\overline{X}(t_n) - \overline{X}(t_{n-1})) - \Delta X(\overline{X}(t_{n-1}))\big) \simeq \hat{e}(t_n), \tag{2.26}$$

motivated by the compensated summation method by Kahan, cf. [14], is useful when the main roundoff error is caused by the recursive summation $\overline{X}(t_{n-1}) + \Delta X(\overline{X}(t_{n-1}))$. Here $\Delta X(\overline{X}(t_{n-1}))$ is the explicit increment $\overline{X}(t_n) - \overline{X}(t_{n-1})$, of the method $\overline{X}$, and fl denotes the rounded operation. A test analogous to (2.25) can be applied also to the dual problem (2.4) for Ψ, to provide an accurate weight function.

2.5 Uniqueness of the Error Representation

To understand a setting for possible uniqueness of the error representation (2.3), suppose that

$$g(X(T)) - g(\overline{X}(T)) = \sum_{i=1}^{N} r_i, \tag{2.27}$$

is an alternative error representation to (2.3). What properties are desirable in order for an error representation to be useful for adaptive mesh refinements? A typical adaptive algorithm does two things iteratively:

(1) if the error indicator is smaller than the given tolerance it stops; otherwise
(2) the algorithm chooses where to refine the mesh and then makes an iterative step to (1).

Therefore the representation r_i must, in addition to estimating the global error (2.27) in (1), also give simple information about where to refine in order to reach the optimal mesh. The only practical method seems to be to link the refinement of element i to the value of r_i. Then an ideal error representation would satisfy:

(i) the error contribution r_i depends only on Δt_i, not on Δt_j for $i \neq j$, and

(ii) $r_i = o(\Delta t_i)$.

The following related conditions imply uniqueness

Theorem 2.5 *Suppose that an error representation satisfies*

$$(i') \quad \text{the error representation (2.27) holds for all choices of step sizes,} \tag{2.28}$$

and

$$(ii') \quad \text{the indicator } r_i \text{ depends only on } \Delta t_k,\ k = 1, \ldots, i, \text{ and not on } \Delta t_j \text{ for } j > i, \tag{2.29}$$

$$(iii') \quad \text{the error indicators have a uniform bound } r_i = o(\Delta t_i). \tag{2.30}$$

Then

$$r_i = \left(e(t_i), \int_0^1 \Psi(t_i, \overline{X}(t_i) + se(t_i))ds \right).$$

Proof. Take the limit $\Delta t_i = 0$, $i = 2, 3, \ldots$. Then by (2.28), (2.30) and Theorem 2.1

$$g(X(T)) - g(\overline{X}(T)) = r_1 = \left(e(t_1), \int_0^1 \Psi(t_1, \overline{X}(t_1) + se(t_1))ds \right). \tag{2.31}$$

Next, let $\Delta t_i = 0$, $i = 3, 4, \ldots$ and use (2.28), (2.29) to get

$$r_1 + r_2 = \sum_{i=1}^{2} \left(e(t_i), \int_0^1 \Psi(t_i, \overline{X}(t_i) + se(t_i))ds \right),$$

which together with (2.31) show also

$$r_2 = \left(e(t_2), \int_0^1 \Psi(t_2, \overline{X}(t_2) + se(t_2))ds \right).$$

Continue this inductive argument to prove the theorem for all r_i. □

Note that the theorem only claims that the indicators r_i are unique: clearly for any invertible $d \times d$ matrix B, we have $(e, \int \Psi ds) = (Be, (B^{-1})^* \int \Psi ds)$. As mentioned, the error representation is not computable and therefore it is not directly useful for adaptive algorithms. Section 3 shows that the computable leading order term of error expansions derived by the variational principle, the error equation and the residual all give the same result, so that in this sense Theorem 2.5 includes them all.

3 Global Error by the Residual and the Error Equation

This section compares the well known derivations of error estimates for differential equations based on the variational principle for the residual [2,11], the classical error equation cf. [13,12] and Galerkin orthogonality for the residual cf. [9,5] to the variational principle for the local error in Theorems 2.1 and 2.2. All methods give the same leading order term, but only the variational principle shows that the local error is a factor in the higher order error terms. The adaptive algorithm in [18] uses only the leading order term. Consequently, computational results with the algorithm in [18] and error indicators based on the three different methods will give the same result.

Below, we use the summation convention, i.e., if the same subscript appears twice in a term, the term denotes the sum over the range of this subscript, e.g.

$$c_{ik}\partial_{x_k} b_j \equiv \sum_{k=1}^{d} c_{ik}\partial_{x_k} b_j.$$

3.1 The Residual with the Variational Principle

The starting point for the residual method is a differential equation

$$\begin{aligned} \frac{dX(t)}{dt} &= a(t, X(t)), \quad t > 0, \\ X(0) &= X_0, \end{aligned} \tag{3.1}$$

and a perturbed equation

$$\begin{aligned} \frac{d\overline{X}(t)}{dt} &= \overline{a}(t, \overline{X}), \quad t > 0, \\ X(0) &= X_0. \end{aligned} \tag{3.2}$$

The error then has the representation

$$g(X(T)) - g(\overline{X}(T)) = \int_0^T (a_i(t, \overline{X}(t)) - \overline{a}_i(t, \overline{X}))\partial_{x_i} u(t, \overline{X}(t))dt, \tag{3.3}$$

where the function $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}$, defined by (2.2), satisfies

$$\begin{aligned} \frac{\partial}{\partial t} u + a_i \partial_{x_i} u &= 0, \quad t < T, \\ u(T, x) &= g(x). \end{aligned} \tag{3.4}$$

The representation (3.3), with $g(x) = x_j$, was derived in [11], by Lie series, and in [2], by using that $\partial_{x_i} u$ is based on the first variation of the solution X, see also [12]. The residual

$$\frac{d\overline{X}}{dt} - a(t, \overline{X}(t)) = \overline{a}(t, \overline{X}) - a(t, \overline{X}(t))$$

shows that (3.3) takes the form error $= \int_0^T$ residual $\times$ weight dt. A derivation of (3.3) based on the transport partial differential equation (3.4) reduces to first observe that by (3.1) and (3.4) the function u is constant along its characteristics

$$\frac{d}{dt} u(t, X(t)) = 0,$$

and in particular

$$u(0, X(0)) = u(T, X(T)). \tag{3.5}$$

Then a consequence of (3.2) and (3.4) is that

$$\begin{aligned} \frac{d}{dt} u(t, \overline{X}) &= \frac{\partial}{\partial t} u(t, \overline{X}(t)) + \overline{a}_i(t, \overline{X}) \partial_{x_i} u(t, \overline{X}(t)) \\ &= (\overline{a}_i(t, \overline{X}) - a_i(t, \overline{X}(t))) \partial_{x_i} u(t, \overline{X}(t)). \end{aligned}$$

Integrate this to obtain

$$u(T, \overline{X}(T)) - u(0, X_0) = \int_0^T (\overline{a}_i(t, \overline{X}) - a_i(t, \overline{X}(t))) \partial_{x_i} u(t, \overline{X}(t)) dt,$$

which combined with (3.5) and the initial data (3.4) shows (3.3).

The close relation between (3.3), error $= \int_0^T$ residual $\times$ weight dt, and (2.3), error $= \sum_n$ local error $\times$ weight, is also explained by

Claim 3.1 *Theorem 2.1 can be derived directly from the representation (3.3).*

Proof. The first step is to determine a numerical flux $\overline{a}$ and to use the representation (3.3) to obtain an error estimate depending on the error of the numerical flux $a - \overline{a}$. Then (3.3) is applied again to replace this residual error by a local discretization error for the solution X.

Step 1. Any numerical method for (3.1) is defined by the nodal values $\overline{X}(t_n)$, $n = 0, 1, \ldots, N$. Every such approximation can be extended to $t \in (0, T)$ by introducing

$$\frac{d\overline{X}}{dt}(t) = \overline{a}(t, \overline{X}), \tag{3.6}$$

for some function $\overline{a}$. The nodal values of $\overline{X}$ only determine

$$\int_{t_{n-1}}^{t_n} \overline{a}(t,\overline{X})dt = \overline{X}(t_n) - \overline{X}(t_{n-1}), \quad n = 1,\dots,N. \tag{3.7}$$

For instance, the forward Euler method has

$$\overline{a}(t,\overline{X}) = a(t_{n-1},\overline{X}(t_{n-1})), \quad t_{n-1} \le t < t_n.$$

Therefore, for a given numerical method of (3.1) we have the freedom to choose any approximate flux $\overline{a}$ satisfying (3.7). We shall use the local error to determine the error of the flux $a - \overline{a}$ and consequently obtain a global error estimate for general methods.

Step 2. The linear backward equation (3.4) shows that

$$u(t_n,\overline{X}(t_n))$$

depends only on the path $\{X\big(s;t_n,\overline{X}(t_n)\big) : \ t_n \le s \le T\}$, so that differentiation gives

$$\partial_{x_i} u(t,\overline{X}(t)) = \frac{\partial u\big(t_n, X(t_n;t,\overline{X}(t))\big)}{\partial X_j(t_n)} \frac{\partial X_j(t_n;t,\overline{X}(t))}{\partial \overline{X}_i(t)}, \tag{3.8}$$

where $\frac{\partial X_j(t_n;t,\overline{X}(t))}{\partial \overline{X}_i(t)}$ denotes the first variation of the exact solution $X(t_n)$ of (3.1) on $[t,t_n]$ with respect to the initial position $X(t) = \overline{X}(t)$ at time t, $t < t_n$, located at the approximate solution. The representation (3.3) applied to the local problem shows

$$\begin{aligned} &X_j\big(t_n;t,\overline{X}(t)\big) - \overline{X}_j(t_n) \\ &\quad = \int_t^{t_n} (a_i(s,\overline{X}(s)) - \overline{a}_i(s,\overline{X})) \frac{\partial X_j(t_n;s,\overline{X}(s))}{\partial \overline{X}_i(s)} ds, \end{aligned} \tag{3.9}$$

so that differentiation of the definition of the local error $e(t)$

$$e(t) \equiv \widetilde{X}(t_n) - X\big(t_n;t,\overline{X}(t)\big), \tag{3.10}$$

where $\widetilde{X}(t_n) = X(t_n;t_{n-1},\overline{X}(t_{n-1}))$ is the exact solution of the local problem (1.7), i.e.

$$\begin{aligned} \frac{d\widetilde{X}}{dt}(t) &= a(t,\widetilde{X}(t)), \quad t_{n-1} < t \le t_n, \\ \widetilde{X}(t_{n-1}) &= \overline{X}(t_{n-1}), \end{aligned} \tag{3.11}$$

implies

$$\frac{de_j}{ds}(s) = (a_i(s,\overline{X}(s)) - \overline{a}_i(s,\overline{X})) \frac{\partial X_j(t_n;s,\overline{X}(s))}{\partial \overline{X}_i(s)}. \tag{3.12}$$

Therefore, the representation (3.3) combined with (3.8), (3.10) and (3.12) establish

$$\begin{aligned}&\int_{t_{n-1}}^{t_n}(a_i(t,\overline{X}(t))-\overline{a}_i(t,\overline{X}))\partial_{x_i}u(t,\overline{X}(t))dt\\&=\int_{t_{n-1}}^{t_n}\partial_{x_j}u(t_n,\widetilde{X}(t_n)-e(t))de_j(t)\\&=u(t_n,\widetilde{X}(t_n))-u(t_n,\overline{X}(t_n)),\end{aligned}$$

which proves the claim. □

3.2 The Error Equation

Let us now compare the estimate (2.15) with the classical estimate obtained from the error equation based on linearization, see [13], neglecting the roundoff error. In [12] the analysis of [13] is generalized to also include the behavior of the higher order error terms, see [12] for more references on such expansions. For a given approximation $\overline{X}$, of (3.1), satisfying (3.6)-(3.7), the first step is to introduce the local approximate solution

$$\begin{aligned}\frac{d\hat{X}}{dt}(t)&=\overline{a}(t,\hat{X}),\quad t_{n-1}<t\le t_n,\\\hat{X}(t_{n-1})&=X(t_{n-1}),\end{aligned}\tag{3.13}$$

which is assumed to be accurate of order $p+1$, i.e.

$$\hat{X}(t_n)-X(t_n)=\mathcal{O}((\Delta t_n)^{p+1}).$$

Then the error $E\equiv\overline{X}-X$ satisfies the error equation

$$\frac{dE}{dt}=\overline{a}(t,\overline{X})-\overline{a}(t,\hat{X})+\overline{a}(t,\hat{X})-a(t,X(t)).$$

Integration of the last two terms yields the local error

$$\int_{t_{n-1}}^{t_n}\big(\overline{a}(t,\hat{X})-a(t,X(t))\big)dt=\hat{X}(t_n)-X(t_n)$$

and a linearization around $\overline{X}$ gives

$$\overline{a}_j(t,\overline{X})-\overline{a}_j(t,\hat{X})=\partial_{x_i}\overline{a}_j(t,\overline{X})(E-\hat{e})_i+\mathcal{O}(|E|^2+|\hat{e}|^2),$$

where $\hat{e} \equiv \hat{X} - X$ is the local error. Let $g(x) \equiv x_k$ and let $\bar{\varphi}$ be the discrete dual function

$$\begin{aligned} -\frac{d\bar{\varphi}_j}{dt} &= \partial_{x_j}\overline{a}_i(t,\overline{X})\bar{\varphi}_i, \quad t < T, \\ \bar{\varphi}_j(T) &= \delta_{jk}, \end{aligned}$$

corresponding to the discrete forward problem

$$\frac{dE_j}{dt} = \partial_{x_i}\overline{a}_j(t,\overline{X})E_i + \frac{d\hat{e}_j}{dt} + \mathcal{O}(|\hat{e}|) + \mathcal{O}(|E|^2 + |\hat{e}|^2).$$

Integration with Duhamel's principle and integration by parts then show that

$$\begin{aligned} &X_k(T) - \overline{X}_k(T) \\ &\quad = -\sum_{n=1}^{N}\int_{t_{n-1}}^{t_n} \frac{d\hat{e}_i(t)}{dt}\bar{\varphi}_i(t)dt \\ &\qquad + \sum_{n=1}^{N}\int_{t_{n-1}}^{t_n} \big(\mathcal{O}((\Delta t_n)^{p+1}) + \mathcal{O}((\Delta t_n)^{2p})\big)|\bar{\varphi}(t)|dt \\ &\quad = \sum_{n=1}^{N}\big(X_i(t_n) - \hat{X}_i(t_n)\big)\bar{\varphi}_i(t_n) \\ &\qquad + \sum_{n=1}^{N}\int_{t_{n-1}}^{t_n} [\big(\mathcal{O}((\Delta t_n)^{p+1}) + \mathcal{O}((\Delta t_n)^{2p})\big)|\bar{\varphi}(t)| \\ &\qquad + \mathcal{O}((\Delta t_n)^{p+1})|\bar{\varphi}'(t)|]dt, \end{aligned} \tag{3.14}$$

where the error terms are based on the local error $\hat{e} = \mathcal{O}((\Delta t_n)^{p+1})$ and the square of the error $|E|^2 = \mathcal{O}((\Delta t_n)^{2p})$. The corresponding a posteriori analysis, based on the local problem (1.7) instead of (3.13), yields similarly

$$\begin{aligned} X_k(T) - \overline{X}_k(T) = &\sum_{n=1}^{N}\big(\widetilde{X}_i(t_n) - \overline{X}_i(t_n)\big)\big(\bar{\varphi}_i(t_n) + \mathcal{O}((\Delta t_n)^p)\big) \\ &+ \sum_{n=1}^{N}\int_{t_{n-1}}^{t_n} \mathcal{O}((\Delta t_n)^{p+1})\big(|\bar{\varphi}(t)| + \mathcal{O}((\Delta t_n)^p)\big)dt. \end{aligned} \tag{3.15}$$

The leading order terms in (3.15) and Theorem 2.2 are the same, while the higher order terms are larger and less precise in the case (3.15) as compared to (2.12), where all terms are directly multiplied by the local error $\widetilde{X}_i(t_n) - \overline{X}_i(t_n)$.

3.3 Galerkin Orthogonality

Our final comparison is with the a posteriori error analysis based on Galerkin orthogonality and residuals in [15,16,10,9,6] and [5]. As a simple example, let the approximation $\check{X}$ be the discontinuous Galerkin approximation of (3.1) with piecewise constants, cf. [9]

$$\check{X}(t_n) = \check{X}(t_{n-1}) + \int_{t_{n-1}}^{t_n} a(s, \check{X}(t_{n-1}))ds \equiv A(\check{X}(t_{n-1})). \quad (3.16)$$

This is a variant of the Euler method with exact quadrature with respect to t in the flux $a(t,x)$ and no roundoff error.

This subsection compares the residual representation (3.3) applied to the discretization (3.16) with the corresponding representation derived by Galerkin orthogonality. Let the approximation $\overline{X}$ of the form (3.2) be defined by $\overline{X}' = \overline{a}(t, \overline{X})$ and

$$\overline{a}(t, \overline{X}) \equiv a(t, \overline{X}(t_{n-1})) \quad t_{n-1} \le t < t_n.$$

Then $\overline{X}$ is continuous and satisfies for all time steps

$$\overline{X}(t_{n-1}) = \check{X}(t_{n-1}).$$

The representation (3.3) implies

$$X_j(T) - \overline{X}_j(T) = \int_0^T (a_k(t, \overline{X}(t)) - \overline{a}_k(t, \overline{X}))\partial_{x_k} u_j(t, \overline{X}(t))dt, \quad (3.17)$$

where $v_{ij} \equiv \partial_{x_i} u_j$ solves

$$\begin{aligned} &-\frac{\partial}{\partial t} v_{ij} - a_k \partial_{x_k} v_{ij} - \partial_{x_i} a_k v_{kj} = 0, \\ &v_{ij}(T, \cdot) = \delta_{ij}. \end{aligned} \quad (3.18)$$

Let $\frac{d}{dt}$ denote differentiation along the characteristic

$$\begin{aligned} &\frac{dX}{ds}(s;t,y) = a(s, X(s;t,y)), \quad s > t \\ &X(t;t,y) = y. \end{aligned} \quad (3.19)$$

Then, $v^y(s) \equiv v(s, X(s))$ is determined by the system of differential equations (3.19) and

$$\begin{aligned} &\frac{d}{ds} v^y(s) = -(a')^T(s, X(s;t,y))\, v^y(s), \ t < s < T \\ &v^y(T) = I, \end{aligned} \quad (3.20)$$

where $(a')^T$ is the transpose of the Jacobian of a and I is the unit matrix in $\mathbb{R}^{d\times d}$. Let us first compare (3.17) with the following well known representation based on local residuals and Galerkin orthogonality, found e.g. in [15], [9], [6].

Claim 3.2 *There holds*

$$X(T) - \check{X}(T) = \int_0^T a(t, \check{X}(t))(\tilde{\varphi}(t) - \Pi\tilde{\varphi}(t))dt, \tag{3.21}$$

where $\tilde{\varphi}(t) = \tilde{\varphi}(t;T)$ solves the linear backward problem,

$$\begin{aligned} &\frac{d\tilde{\varphi}}{dt} = -\alpha(t)\tilde{\varphi}(t), \quad \tilde{\varphi}(T) = I, \\ &\alpha(t) = \int_0^1 (a')^T(t, sX(t) + (1-s)\check{X}(t))ds, \end{aligned} \tag{3.22}$$

(which depends on both the exact solution X and the approximate solution $\check{X}$) and $\Pi\tilde{\varphi}$ is the piecewise constant function

$$\Pi\tilde{\varphi}(t) = \tilde{\varphi}(t_n), \quad \forall t \in (t_{n-1}, t_n].$$

A proof of the Claim is included in the end of this section. Equation (3.22) yields

$$\tilde{\varphi}(t) - \Pi\tilde{\varphi}(t) = -\int_{t_n}^t \alpha(s)\tilde{\varphi}(s)ds = (\alpha\tilde{\varphi})(\xi_n(t))(t_n - t)$$

for some $\xi_n(t) \in (t, t_n)$. Therefore the representation (3.21) takes the form

$$X(T) - \check{X}(T) = \sum_{n=1}^N \int_{t_{n-1}}^{t_n} a(t, \check{X}(t)) \cdot (\alpha\tilde{\varphi})(\xi_n(t))(t_n - t)dt. \tag{3.23}$$

On the other hand, the estimate (3.17) can be written

$$\begin{aligned} X(T) - \check{X}(T) = \sum_{n=1}^N \int_{t_{n-1}}^{t_n} \Big\{ \int_0^1 a'(t, \overline{X}(t_{n-1})s + \overline{X}(t)(1-s))ds \Big\} \\ \times \overline{a}(t, \overline{X})(t - t_{n-1}) \cdot v^{\overline{X}(t)}(t)dt, \end{aligned} \tag{3.24}$$

which is clearly similar to (3.23). One difference is that in (3.23), the matrix $\alpha^T(\xi_n)$ is the averaged Jacobian of a' between the exact and approximate solutions $X(t)$ and $\check{X}(t)$, with difference $X(t) - \check{X}(t) = \mathcal{O}(\Delta t)$; while in (3.24) we have the less averaged Jacobian of a' between the approximate solution at time t and at t_{n-1}, with smaller

difference $\overline{X}(t)-\overline{X}(t_{n-1})=\mathcal{O}((\Delta t)^2)$. A second difference, is that in (3.23) the differential equation for $\tilde{\varphi}$ has a linear flux function based on the averaged Jacobian α, while in (3.24) the function $v^{\overline{X}(t)}(t)$ is the solution of the differential equation (3.19)-(3.20) with the linear flux based in the Jacobian $a'(s,X^{\overline{X}(t)}(s))$ along exact solutions, starting at the approximate solution $\overline{X}(t)$.

The work [5] derives a different error representation, also based on Galerkin orthogonality, which for (3.16) leads to a similar error representation

$$\begin{aligned}X(T)-\check{X}(T)=\frac{1}{2}\sum_{n=1}^{N}\int_{t_{n-1}}^{t_n}\check{a}(t)\cdot\int_t^{t_n}(a'(s))^T v^{X(s)}(s)dsdt\\ +\frac{1}{2}\sum_{n=1}^{N}\int_{t_{n-1}}^{t}\Big(a(s)ds\cdot(\check{a}'(t))^T\varphi(t)+R(t)\Big)dt\end{aligned}$$

where

$$\begin{aligned}&R=\mathcal{O}((|X(t)-\check{X}(t)|+|v^{X(t)}(t)-\bar{\varphi}(t)|)^3),\\ &\check{a}(s)\equiv a(s,\check{X}(s)),\ \check{a}'(s)\equiv a'(s,\check{X}(s)),\end{aligned}$$

and

$$\bar{\varphi}_i(t_n)=\partial_{x_i}A(\check{X}(t_n))\bar{\varphi}(t_{n+1}),\quad \bar{\varphi}_i(T)=\partial_{x_i}g(\check{X}(T)).$$

Proof. Proof of Claim 3.2, as in [9,16]. Let $\epsilon\equiv X-\check{X}$. Then

$$\begin{aligned}L\epsilon\equiv\epsilon'-\alpha^T\epsilon&=X'-a(\cdot,X)-(\check{X}'-a(\cdot,\check{X}))\\ &=-(\check{X}'-a(\cdot,\check{X}))\equiv R,\end{aligned}\tag{3.25}$$

where the local residual R is a bounded measure on $[0,T]$. The construction (3.22) can then be written

$$L^*\tilde{\varphi}=0.\tag{3.26}$$

Let $(v,w)\equiv\int_0^T v(s)\cdot w(s)ds$ and integrate by parts to obtain

$$(L\epsilon,\tilde{\varphi})=(\epsilon,L^*\tilde{\varphi})+X(T)-\check{X}(T).\tag{3.27}$$

On the other hand, there holds

$$(L\epsilon,\tilde{\varphi})=(R,\tilde{\varphi}),\tag{3.28}$$

and the Galerkin orthogonality applied to (3.16) yields

$$(R,\tilde{\varphi})=(R,\tilde{\varphi}-\Pi\tilde{\varphi})=(-a(t,\check{X}(t)),\tilde{\varphi}-\Pi\tilde{\varphi}),\tag{3.29}$$

since $\tilde{\varphi}(t_{n-1})-\Pi\tilde{\varphi}(t_{n-1})=0$. Together (3.25)-(3.29) prove (3.21). □

Acknowledgements This work has been supported by the EU–TMR project HCL # ERBFMRXCT960033, the EU–TMR grant # ERBFMRX-CT98-0234 (Viscosity Solutions and their Applications), the Swedish Science Foundation, UdelaR and UdeM in Uruguay, the Swedish Network for Applied Mathematics, the Parallel and Scientific Computing Institute (PSCI) and the Swedish National Board for Industrial and Technical Development (NUTEK).

References

1. M. Ainsworth and J. T. Oden, A posteriori error estimation in finite element analysis, Comput. Methods Appl. Mech. Engrg., **142** (1997), 1-88.
2. V.M. Alekseev, An estimate for the perturbations of the solutions of ordinary differential equations. II, Vestnik Moskov. Univ. Ser. I Mat. Mech, **3** (1961), 3-10.
3. I. Babuška, A. Miller and M. Vogelius, Adaptive methods and error estimation for elliptic problems of structural mechanics, in *Adaptive computational methods for partial differential equations* (SIAM, Philadelphia, Pa., 1983) 57-73.
4. R. Becker and R. Rannacher, A feed-back approach to error control in finite element methods: basic analysis and examples, East-West J. Numer. Math., **4** (1996), no. 4, 237-264.
5. R. Becker and R. Rannacher, An optimal control approach to a posteriori error estimation in finite element methods Acta Numerica, (2001), 1-102.
6. K. Böttcher and R. Rannacher, Adaptive error control in solving ordinary differential equations by the discontinuous Galerkin method, preprint, (1996).
7. G. Dahlquist and Å. Björk, *Numerical Methods*, (Prentice-Hall, 1974).
8. G. Dahlquist and Å. Björk, *Numerical Mathematics*, http://www.mai.liu.se/∼akbjo/NMbook.html.
9. K. Eriksson, D. Estep, P. Hansbo and C. Johnson, Introduction to adaptive methods for differential equations, Acta Numerica, (1995), 105-158.
10. D. Estep, A posteriori error bounds and global error control for approximation of ordinary differential equations, SIAM J. Numer. Anal., **32** (1995), 1-48.
11. W. Gröbner, *Die Lie-Reihen und ihre Anwendungen*, (VEB Deutscher Verlag der Wissenschaften, Berlin, 1967).
12. E. Harrier, S.P. Norsett and G. Wanner, *Solving Ordinary Differential Equations I*, (Springer-Verlag, 1993).
13. P. Henrici, *Discrete Variable Methods in Ordinary Differential Equations* (John Wiley & Sons, Inc., 1962).
14. N.J. Higham, *Accuracy and Stability of Numerical Algorithms*, (Society for Industrial and Applied Mathematics, 1996).
15. C. Johnson, Error estimates and adaptive time-step control for a class of one-step methods for stiff ordinary differential equations, SIAM J. Numer. Anal., **25** (1988), 908-926.
16. C. Johnson and A. Szepessy, Adaptive finite element methods for conservation laws based on a posteriori error estimates, Comm. Pure Appl. Math., **48** (1995), 199-234.
17. K.-S. Moon, A. Szepessy, R. Tempone and G.E. Zouraris, Hyperbolic differential equations and adaptive numerics, in *Theory and numerics of differential equations* (Eds. J.F. Blowey, J.P. Coleman and A.W. Craig, Durham 2000, Springer Verlag, 2001).

18. K.-S. Moon, A. Szepessy, R. Tempone and G.E. Zouraris, Convergence rates for adaptive approximation of ordinary differential equations, Numer. Math. (in press).
19. A. Prothero, Estimating the accuracy of numerical solutions to ordinary differential equations, in *Computational Techniques for Ordinary Differential Equations*, I. Gladwell and D.K. Sayers, eds. Academic Press, New York, 1980, 103-128.
20. T. Utumi, R. Takaki and T. Kawai, Optimal time step control for the numerical solution of ordinary differential equations, SIAM J. Numer. Anal., **33** (1996), 1644-1653.